

Nonanimal Models for Acute Toxicity Evaluations: Applying Data-Driven Profiling and Read-Across

Daniel P. Russo,¹ Judy Strickland,² Agnes L. Karmaus,² Wenyi Wang,¹ Sunil Shende,^{1,3} Thomas Hartung,^{4,5} Lauren M. Aleksunes,⁶ and Hao Zhu^{1,7}

¹Center for Computational and Integrative Biology, Rutgers University, Camden, New Jersey, USA

²Integrated Laboratory Systems (ILS), Research Triangle Park, North Carolina, USA

³Department of Computer Science, Rutgers University, Camden, New Jersey, USA

⁴Johns Hopkins Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), Baltimore, Maryland, USA

⁵University of Konstanz, CAAT-Europe, Konstanz, Germany

⁶Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey, USA

⁷Department of Chemistry, Rutgers University, Camden, New Jersey, USA

BACKGROUND: Low-cost, high-throughput *in vitro* bioassays have potential as alternatives to animal models for toxicity testing. However, incorporating *in vitro* bioassays into chemical toxicity evaluations such as read-across requires significant data curation and analysis based on knowledge of relevant toxicity mechanisms, lowering the enthusiasm of using the massive amount of unstructured public data.

OBJECTIVE: We aimed to develop a computational method to automatically extract useful bioassay data from a public repository (i.e., PubChem) and assess its ability to predict animal toxicity using a novel bioprofile-based read-across approach.

METHODS: A training database containing 7,385 compounds with diverse rat acute oral toxicity data was searched against PubChem to establish *in vitro* bioprofiles. Using a novel subspace clustering algorithm, bioassay groups that may inform on relevant toxicity mechanisms underlying acute oral toxicity were identified. These bioassays groups were used to predict animal acute oral toxicity using read-across through a cross-validation process. Finally, an external test set of over 600 new compounds was used to validate the resulting model predictivity.

RESULTS: Several bioassay clusters showed high predictivity for acute oral toxicity (positive prediction rates range from 62–100%) through cross-validation. After incorporating individual clusters into an ensemble model, chemical toxicants in the external test set were evaluated for putative acute toxicity (positive prediction rate equal to 76%). Additionally, chemical fragment–*in vitro*–*in vivo* relationships were identified to illustrate new animal toxicity mechanisms.

CONCLUSIONS: The *in vitro* bioassay data-driven profiling strategy developed in this study meets the urgent needs of computational toxicology in the current big data era and can be extended to develop predictive models for other complex toxicity end points. <https://doi.org/10.1289/EHP3614>

Introduction

There are currently over 100,000 chemicals available on the market that lack toxicity information, comprising roughly 90% of the 140,000 consumer products in use (Hartung and Rovida 2009; Judson et al. 2009). Traditional toxicology evaluations require the use of animal models for testing new compounds. However, these animal models are costly and time-consuming, and they raise ethical concerns regarding the well-being of animals (Hartung 2017). Under this paradigm, generating substantial toxicity data for a limited number of compounds could take years, and it would be financially impossible to test all the available compounds using animal testing protocols (Hartung 2016). In 2007, the National Research Council Committee on Toxicity Testing and Assessment of Environmental Agents addressed this issue by proposing a new framework to accurately and more quickly evaluate the health risks due to environmental chemical exposures (National Research Council 2007). This federal effort stressed the importance of integrating/establishing the use of computational and *in vitro*–based alternative methods for chemical risk evaluation. One such alternative, called read-across,

relies on using toxicity information from structurally similar compounds to estimate the toxicity of untested compounds (Patlewicz et al. 2014; Wang et al. 2012). This strategy can be used to fill toxicity data gaps for untested chemicals and has been implemented by various regulatory agencies (Hartung 2016). Previous read-across studies relied solely on chemical structure similarity searching (Enoch et al. 2008; Hewitt et al. 2010; Koleva et al. 2008; Luechtefeld et al. 2016; Wu et al. 2013). However, this type of read-across is not applicable for compounds with unique chemical structures and can be confounded by “activity cliffs” (i.e., structurally similar compounds with distinctly different toxicity characteristics) (Maggiore 2006). More recently, efforts to include biological information as a basis for similarity in read-across approaches have started (Zhu et al. 2016). Previous studies using biological data for chemical toxicity evaluations were mostly based on in-house biological data and were limited to specific mechanisms (Judson et al. 2015; Kleinstreuer et al. 2017). This paper addresses the challenge of identifying and integrating biological data from various resources into read-across modeling.

In vitro high-throughput screening (HTS) is capable of rapidly testing large numbers of chemicals to study their effects on molecular targets using whole-cell and cell-free assays. Because of their relatively low cost and high-throughput, efforts such as the Toxicity Testing in the 21st Century (Tox21) program have focused on the application of HTS techniques as the basis for chemical hazard assessment (Attene-Ramos et al. 2013). The direct result of these efforts is a rapidly growing amount of *in vitro* bioassay data being generated for thousands of chemicals and stored in databases accessible to public users, allowing for new statistical and computational techniques to be developed. The impact of such large publicly available databases for chemical toxicity evaluation is profound, with several projects having successfully used HTS data to better evaluate chemicals for potential hazards (Browne et al. 2015; Hartung 2016; Kim et al. 2016; Kleinstreuer et al.

Address correspondence to Hao Zhu, 315 Penn St., Rutgers University, Camden, NJ 08102, USA. Telephone: (856) 225-6781. Email: hao.zhu99@rutgers.edu

Supplemental Material is available online (<https://doi.org/10.1289/EHP3614>). The authors declare they have no actual or potential competing financial interests.

Received 9 March 2018; Revised 6 March 2019; Accepted 8 March 2019; Published 1 April 2019.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

2017; Low et al. 2013; Zhang et al. 2014a; Zhu et al. 2014). However, rapidly changing public data sources represent a dynamic data landscape, and integrating such data to chemical information for toxicity evaluation is an area that remains largely unexplored. Development of automated computational methods to deeply exploit this rich and dynamic data landscape to establish predictive nonanimal toxicity models is needed.

Acute oral toxicity testing is conducted to determine the immediate health effects of an orally administered chemical substance and is expressed in terms of the lethal dosage that kills 50% of the population (LD_{50}) of animals tested (Strickland et al. 2018). Acute oral toxicity data are used by a number of regulatory agencies for hazard classification and labeling of products to alert handlers and consumers of potential toxicity hazards, to determine acceptable human exposure limits and personal protective equipment needed for handling, and determine countermeasures that should be employed in the event of toxic exposures (Corvaro et al. 2016; Strickland et al. 2018; Walum 1998). In some cases, acute oral toxicity data may be used to establish doses for longer-term studies, identify target organs for toxicity, and assess the hazard of accidental ingestions of chemical contaminants in food (Strickland et al. 2018). To date, there are no *in vitro* tests accepted by regulatory agencies as stand-alone replacements for acute oral animal tests (Kinsner-Ovaskainen et al. 2009; Strickland et al. 2018).

Here, we present a new computational technique to automatically extract pertinent data from PubChem, which is updated daily, and develop a predictive ensemble model for estimating acute oral toxicity (outlined in Figure 1). In this work, a large dataset consisting of compounds with broad acute oral toxicity values was profiled by their PubChem *in vitro* bioassay data to

generate bioprofiles. Using these initial bioprofiles, we characterized and clustered mechanistically similar PubChem bioassays using fragments of the chemicals tested within them. The resulting chemical fragment-*in vitro*-*in vivo* relationships were the foundation for bioprofile-based read-across studies conducted using clusters of PubChem bioassays considered to inform on similar toxicity mechanisms.

Methods

Acute Oral Toxicity Datasets

An in-house rat acute oral toxicity database was previously collected and curated from ChemIDplus (<https://chem.nlm.nih.gov/chemidplus/>; Zhu et al. 2009). This in-house rat acute oral toxicity database was used as the training set and consisted of 7,385 compounds with their most conservative LD_{50} values (i.e., lowest recorded LD_{50} value for a single compound) ranging from 0.02 $\mu\text{g}/\text{kg}$ to 389.39 g/kg (Figure 2A). For modeling purposes, the LD_{50} values were normalized by the following logistic function:

$$f(x) = \frac{e^{\frac{x-x_0}{s}}}{1 + e^{\frac{x-x_0}{s}}}, \quad (1)$$

where x represents a log10-transformed LD_{50} ($\log LD_{50}$) value, x_0 represents the midpoint of the curve, and s is a number to control the shape of the curve. Figure 2B shows the relationship between the logistic function outputs and the $\log LD_{50}$ values through various values of s . The outputs of the logistic function are all between 0 and 1, so we chose the threshold of 0.5 to distinguish toxic ($f \geq 0.5$) and nontoxic ($f < 0.5$) compounds. To obtain a

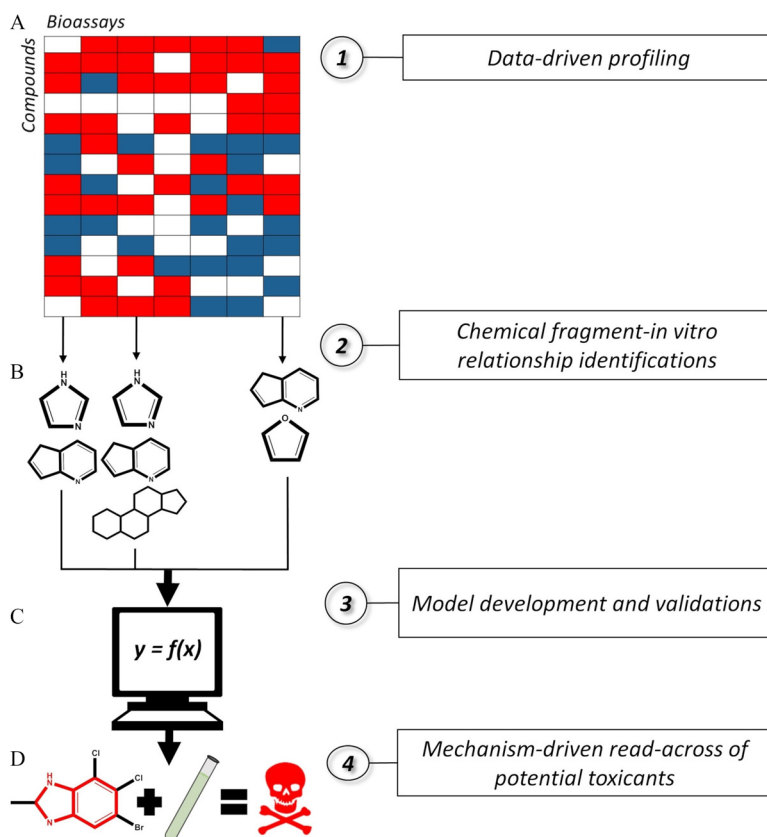


Figure 1. Overview of the data-driven read-across approach developed in this study. (A) Profiling compounds via PubChem portal, (B) clustering PubChem assays based on chemical fragment *in vitro* relationships, (C) identifying bioassays capable of predicting acute oral toxicity, and (D) predicting new toxicants by read-across and mechanism illustrations.

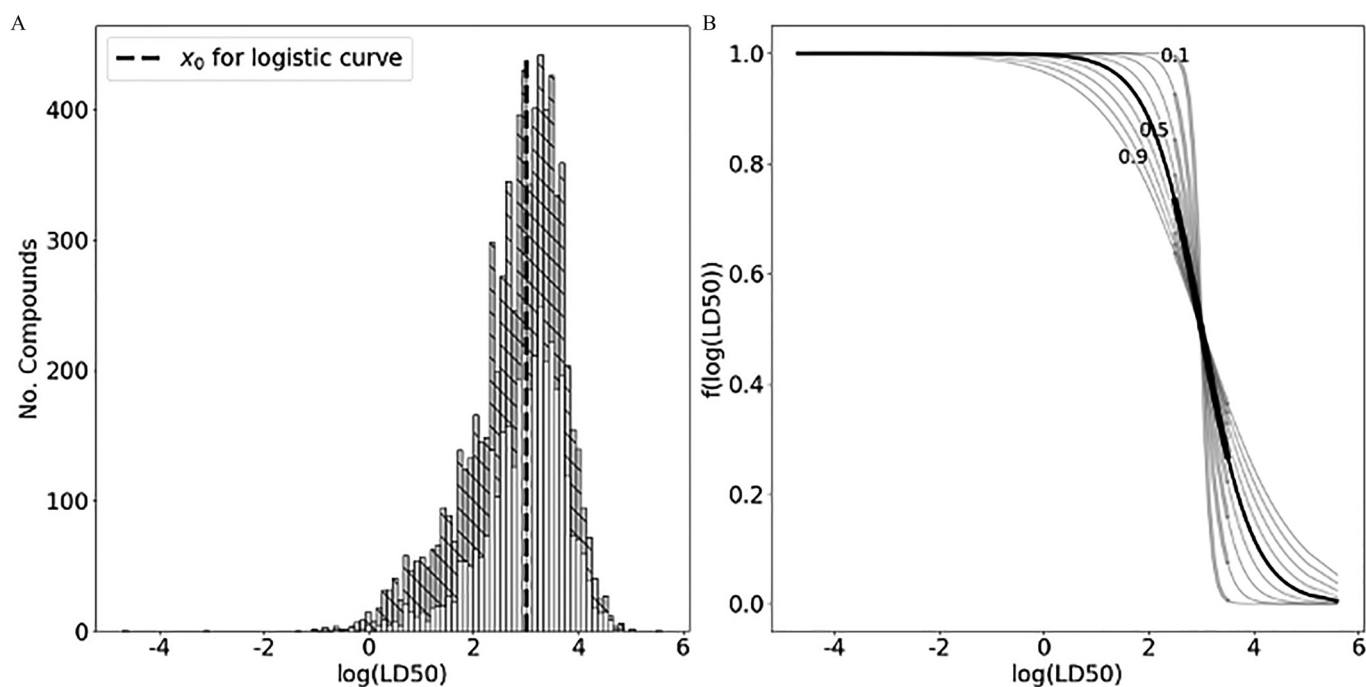


Figure 2. Activity distribution for compounds in this study. (A) Histogram for the 7,385 log-transformed LD₅₀ values from compounds in the original dataset (hashed bars in background) and the final 3,543 compounds with biological data ultimately used for modeling (solid bars in foreground). The parameter, x_0 , is equal to the midpoint of the logistic curve displayed in (B) and was used as the cutoff for determining compound toxicity. (B) The parameter s controls the shape of the curve. Lines represent the logistic function at various values for s (0.1 through 0.9 at steps of 0.1). The final curve, with $s=0.5$, is bold. The thin areas on the line indicate log-transformed LD₅₀ values <2.5 or >3.5 and their corresponding values in f space. The thick line indicate log-transformed LD₅₀ values ≥ 2.5 and ≤ 3.5 .

balanced dataset and easily interpretable cutoff, we set the midpoint (x_0) at 3 (which is equivalent to 1,000 mg/kg), yielding 3,791 toxic and 3,594 nontoxic compounds.

A separate dataset consisting of 3,852 compounds with rat acute oral LD₅₀ values was collected from a variety of sources, including ChemIDplus (<https://chem.nlm.nih.gov/chemidplus/>), the Hazardous Substances Data Bank (<https://toxnet.nlm.nih.gov/newtoxnet/hsdb.htm>), European Chemical Agency (<https://echa.europa.eu/information-on-chemicals>), and the U.S. Environmental Protection Agency (EPA) (U.S. EPA 2016). This dataset served as an external test set to evaluate the generated models. We refined this dataset by excluding compounds already included in the training set, standardizing chemical structures, and removing compounds with an LD₅₀ value reported as a range (e.g., LD₅₀ > 500 mg/kg). Thus, the curated external test set ultimately contained 639 compounds with LD₅₀ values ranging from 0.012 mg/kg to 63.97 g/kg. These LD₅₀ values were also converted to classifications using the logistic function (including the same threshold and midpoint) as described above.

Bioprofiling and Subspace Clustering of PubChem Data

To form a training set of compounds for modeling, public *in vitro* bioassay data for 7,385 compounds were extracted from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) using an automatic data mining portal (<http://ciiipro.rutgers.edu/>) (Russo et al. 2017; Zhang et al. 2014a). To establish meaningful relationships between chemical fragment descriptors and PubChem bioassays, a feature reduction approach was applied. Briefly, those bioassays with very limited data across the chemicals in our training set were removed to avoid overfitting by training the model using minimal signal. Thus, only bioassays with at least five active responses among the training set compounds were included in the modeling procedure. This effort resulted in a sufficiently large dataset containing 3,543 training set compounds with *in vitro* data from 1,077 PubChem bioassays.

The bioactivity values of the 3,543 training set compounds across each of the 1,077 PubChem assays comprised the initial bioprofile. To identify potential toxicity mechanisms and further optimize the initial bioprofile, the 1,077 PubChem bioassays were clustered based on shared chemical fragments relevant to bioassay responses. To achieve this, we used the established ToxPrint fingerprints, a set of 729 chemical fragments relevant to toxicity reported in a previous study (Yang et al. 2015); fingerprints were generated using ChemoTyper (Molecular Networks GmbH, Erlangen, Germany) software (version 1.0). Then, the relationship between these ToxPrint fragments and each PubChem bioassay were determined using Fisher's exact test. Fisher's exact test requires constructing a 2×2 contingency matrix, which we define below:

	Compounds with fragment	Compounds without fragment
Compounds active in assay	a	b
Compounds inactive in assay	c	d

(2)

where a is the number of compounds with an active response in this assay and contain this fragment, b is the number of compounds with an active response and do not contain the fragment, c is the number of compounds with an inactive response and contain the fragment, and d is the number of compounds with an inactive response and do not contain the fragment.

The output of this test is a p -value denoting the statistical significance of the relationship between the fragment and bioassay activity. When considering multivariate comparisons, it is necessary to ensure that the resulting correlations are not spurious. However, strict multiple testing corrections, such as the Bonferroni correction, greatly reduced the number of significant relationships (<90% and a loss of ~60% of the biological data) and were conservative for inclusion purposes. Alternatively, the number of false positives was approximated in this analysis by comparing p -values to those obtained from a permutation test (Figure S1). The standard p -value

threshold (p -value < 0.05) was chosen to define significant relationships, which was sufficient for minimizing spurious correlations.

PubChem bioassays sharing many significant fragments could be related and/or unveil potential mechanisms of oral acute toxicity for specific chemical toxicants. To group similar assays, the Jaccard dissimilarity (J_d) between each bioassay was calculated using the fragment profile, defined as:

$$J_d(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}, \quad (3)$$

where A and B represent the sets of significant fragments for PubChem bioassays A and B , respectively. Calculating dissimilarity (J_d) between bioassays allows for the representation of potential relationships among assays as a network graph, where nodes represent bioassays, and edges are the J_d values between two bioassays. The network graph was created and manipulated using the software package Gephi (<https://gephi.org/>) (version 0.9.1). Clusters of PubChem assays were determined by using the Louvain modularity algorithm available within the Gephi software package, using the “resolution” parameter to determine the bioassays that belong to the same cluster (Blondel et al. 2008). A larger resolution value allows more bioassays to form larger clusters, and a smaller resolution value restricts bioassays with higher similarity to small clusters. A resolution value of 0.25 was used in this study, with the maximum number of bioassays in the resulting clusters set to 60.

Bioassay-Based Read-Across for Acute Oral Toxicity Classifications

We delineated clusters of bioassays that were highly predictive of acute toxicity to identify relevant bioassays and unveil potential toxicity mechanisms. Read-across studies were then performed using the bioassay responses within a cluster to predict acute oral toxicity classification; prediction results were evaluated by five-fold cross-validation. If bioassays showed highly correlated responses with at least 10 mutual test responses ($R^2 > 0.9$), one of them was randomly selected and incorporated into model building, and the other highly correlated assays were removed from any further analysis. Training set compounds with at least one *in vitro* bioassay response in a cluster were randomly split into five equivalent subsets. For each of five iterations, one subset acted as a pseudotest set, while the remaining compounds served as a pseudotraining set. The acute oral toxicity classification of each validation set compound was predicted by the most biosimilar compound in the modeling set, based on sharing the most similar PubChem bioassay responses within the cluster. The biosimilarity between two compounds in a cluster c can be calculated by:

$$\text{Biosimilarity}(A, B)_c = \frac{|A_a \cap B_a| + |A_i \cap B_i| \cdot w}{|A_a \cap B_a| + |A_i \cap B_i| \cdot w + |A_a \cap B_i| + |A_i \cap B_a|}, \quad (4)$$

where A_a and B_a represent the sets of active responses in PubChem bioassays within a cluster c for compounds A and B , respectively. Conversely, the terms A_i and B_i represent the sets of inactive responses. The term w weights the inactive responses less than active responses since the proportion of active data, which indicates more significant biological phenomena, is much lower than inactive data. In this study, w was calculated as the ratio of total active responses to total inactive responses for each cluster.

Public bioassay data is inherently sparse, and therefore, calculating biosimilarity alone can offer misleading results due to

missing data. For example, when two compounds are both only tested in one assay within a cluster, their biosimilarity result is less reliable than two other compounds that share responses in multiple assays. For this reason, the relative cluster confidence (rcc) of a biosimilarity calculation was also computed using the equation below:

$$rcc(A, B)_c = \frac{|A_a \cap B_a| + |A_i \cap B_i| \cdot w + |A_a \cap B_i| + |A_i \cap B_a|}{N}, \quad (5)$$

where N is the number of noncorrelated assays in cluster c (i.e., total number of bioassays used in the model). Here, a high rcc is indicative of the two compounds being tested in many of the same assays within a cluster.

The underlying mechanisms responsible for toxicity in animal acute oral studies are vast, and thus, it is unlikely to expect a few bioassays to explain all the various toxic phenomena. Therefore, when using bioassays as models for acute oral toxicity, it is reasonable to expect a relatively high false negative rate (compounds inactive in bioassay response yet active in acute oral animal toxicity test). In these cases, toxicity may be elicited via mechanisms that are not represented by the biochemical coverage of the *in vitro* assays mined. Therefore, to focus on the ability to identify toxic compounds, the positive predictive value (ppv) was used to evaluate the model performance and is defined as

$$ppv = \frac{TP}{TP + FP}, \quad (6)$$

where TP represents the number of true positives (toxic compounds correctly predicted as toxic), and FP represents the number of false positives (nontoxic compounds incorrectly predicted as toxic).

Quantitative Structure–Activity Relationship Models

Missing data severely limits the identification of relationships in this study. Simple data imputation methods, such as random sampling, are not sufficiently robust and may create further issues. In order to correctly impute biological data, especially for the toxicants with missing data, more reasonable data imputation methods were necessary. To realize this, Random Forest and Naïve Bayes, both implemented using the Python library scikit-learn (version 0.18.1), were used to develop quantitative structure–activity relationship (QSAR) models (Pedregosa et al. 2011). These QSAR approaches were implemented for PubChem bioassays to fill in the missing bioassay data, allowing for sufficient data points to evaluate chemical fragment–*in vitro*–*in vivo* relationships (i.e., when prediction performance is increased by the presence of a toxicophore). More specifically, of the original 7,385 chemicals identified from the in-house rat acute oral toxicity database, only 3,543 had sufficient bioassay data for initial analyses. To increase scope for later chemical fragment–*in vitro*–*in vivo* relationship analyses, the remaining 3,842 chemicals were subjected to the QSAR workflow.

Random Forest is an ensemble algorithm consisting of constructing many decision trees and then making a prediction by combining the output among the trees (Breiman 2001). Naïve Bayes is an algorithm that predicts by estimating the probability of membership to a certain class using Bayes theorem (Friedman et al. 1997). The QSAR model development herein followed the workflow used in our previous studies and briefly outlined below (Kim et al. 2014; Solimeo et al. 2012; Sprague et al. 2014; Wang et al. 2015; Zhao et al. 2017).

The training data for QSAR model development was retrieved from PubChem’s PUG-REST web service by using a PubChem

Assay Identifier (AID) as the query (Kim et al. 2015). The information retrieved for a single bioassay included chemical structures and activity classifications (active/inactive/inconclusive) for all tested compounds in the bioassay. First, inconclusive results were eliminated. Then, the active/inactive ratio was balanced by randomly selecting and eliminating compounds until an equal ratio was achieved. Training set data was set to not exceed 10,000 compounds. The rdkit implementation of extended-connectivity fingerprints was used as chemical features for the remaining compounds in QSAR model training (Rogers and Hahn 2010).

A fivefold cross-validation procedure available within the scikit-learn package was used to evaluate the resulting models. This procedure searched and stored the models based on fivefold cross-validation prediction accuracy (*acc*) as defined as:

$$acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (7)$$

where true positive (*TP*) is the number active compounds correctly predicted as active, false positive (*FP*) is the number of inactive compounds incorrectly predicted as active, false negative (*FN*) is the number of active compounds incorrectly predicted as inactive, and true negative (*TN*) is the number of inactive compounds correctly predicted as inactive. Only models with a cutoff of *acc* > 0.6 were used to fill data gaps for substances with missing *in vitro* bioassay results. If both Random Forest and Naïve Bayes models had an *acc* > 0.6 for the same PubChem assay, the model with higher *acc* value was used.

Mechanism-Driven Toxicity Pathway Analysis

By integrating QSAR predictions into bioprofiles to address missing data (i.e., for chemicals that were not tested), chemical fragment-*in vitro-in vivo* relationship analyses can be performed by examining the chemical fragments of the compounds with predicted bioassay results. If compounds contain a specific chemical fragment that produces superior prediction accuracy within a cluster, this chemical fragment can be selected as a potential toxicophore, which is the principal chemical feature to induce a toxicity pathway (Allen et al. 2014). In this way, new toxicity mechanisms can be revealed by linking chemical fragments, relevant *in vitro* bioprofiles, and *in vivo* acute oral toxicity.

Results

Acute Oral Toxicity Classifications

A logistic function (*f*) was used in this study to convert experimental log-transformed LD₅₀ (logLD₅₀) values into classifications (toxic/nontoxic), applying a threshold of *f* ≥ 0.5 to define toxic compounds based on an established LD₅₀ cutoff of 1,000 mg/kg. This threshold balances the toxic/nontoxic ratio (see “Methods” section) and is close to the EPA’s current criterion to classify acute oral toxicants as Category II (50 < LD₅₀ ≤ 500 mg/kg) or Category III (500 < LD₅₀ ≤ 5,000 mg/kg) (U.S. EPA 2012).

Many compounds have logLD₅₀ values close to the threshold (i.e., a logLD₅₀ of 3 ± 0.5), making these compounds difficult to classify as overtly toxic or nontoxic. Changing the parameter *s* has the effect of shrinking or extending the range of the outputted values of compounds close to this threshold (bold line in Figure 2B). By selecting an *s* value of 0.5, the outputted *f* value of these compounds would fall between 0.75 and 0.25 (e.g., logLD₅₀ value as 2.5 with the associated *f* output as 0.75). As shown in Figure 2A,B, many compounds in the database can fall within these two areas.

Subspace Clustering of PubChem Assays

Among the 1,077 PubChem bioassays in the original bioprofile, 707 had at least one ToxPrint chemical fragment that was significantly correlated with activity, resulting in a total of 15,064 significant chemical fragment-*in vitro* bioassay activity relationships (*p*-value < 0.05). These relationships were used to cluster the PubChem assays by calculating the dissimilarity (*J_d*) between every pair of bioassays. As shown in Figure 3, the edges between two nodes (bioassays) indicate dissimilarity values less than 0.75. Sixty-seven bioassays are unique compared to others (no *J_d* values less than 0.75) and are not shown in this figure. Within the remaining 640 bioassays, the Louvain modularity algorithm (Blondel et al. 2008) identified 45 unique clusters with 2 to 60 bioassays per cluster (Excel Table S1). The clusters result in grouping bioassays that are potentially related in their ability to inform on biological mechanism as it pertains to acute oral toxicity.

The 640 bioassays came from different source depositors, which are summarized in Table 1. Bioassays from different sources coexisted within clusters, with 37 out of 45 clusters containing bioassays from ≥ two sources. The biological targets of bioassays varied, and the majority could be classified as overt toxicity (e.g., cytotoxicity assays), biomarkers of cellular responses (e.g., mitochondrial membrane potential assays), or specific protein targets (e.g., agonists of the androgen receptor). Some clusters showed a clear preponderance of one bioassay type (e.g., Cluster 1 consisted of only cytotoxicity assays).

Acute Oral Toxicity Model Selection and Ensemble Modeling

For modeling purposes, the 45 PubChem bioassay clusters shown in Figure 3 were evaluated for their ability to predict acute oral toxicity. Bioprofile-based read-across studies were performed within each cluster and assessed using a fivefold cross-validation procedure. Different thresholds of biosimilarity (ranging from 0.5 to 0.9) and *rcc* values (ranging from 0 to 100%) were used to determine the most similar compound in the training set to a test compound. The best *ppv* for each cluster was recorded, with cross-validated *ppv* values from 0 to 100% across the bioassay clusters (Figure 4). The clusters with a *ppv* above 60% were selected as viable potential models of acute oral toxicity. To ensure the models are statistically significant and interpretable, clusters with < five bioassays were removed. This procedure resulted in 19 PubChem bioassay models that were potentially applicable for acute oral toxicity prediction. In order to leverage predictions across multiple models (and thus multiple potential biological mechanisms), an ensemble model was created by averaging the predictions across all the cluster-specific models.

PubChem Bioassays Selected for Acute Toxicity Classifications

The PubChem bioassays identified by our approach for toxicity predictions came from a variety of screening programs. These programs have different research goals and vary in scope and size, and many were not designed to be used for acute oral toxicity evaluation purposes. However, the analysis of the top-ranked cluster-based models unveiled several mechanisms clearly relevant to acute oral toxicity, including cytotoxicity/growth inhibition in cells of various origins including tumor cell lines, viral growth inhibition, and assays measuring protein interaction and/or function.

Cluster 1 had the highest accuracy in the cross-validation procedure (100% *ppv*). It consisted entirely of data deposited from the Developmental Therapeutics Program at the National Cancer Institute (NCI). The 17 NCI bioassays comprising Cluster 1 measure cytotoxicity/growth inhibition against tumor cell lines and

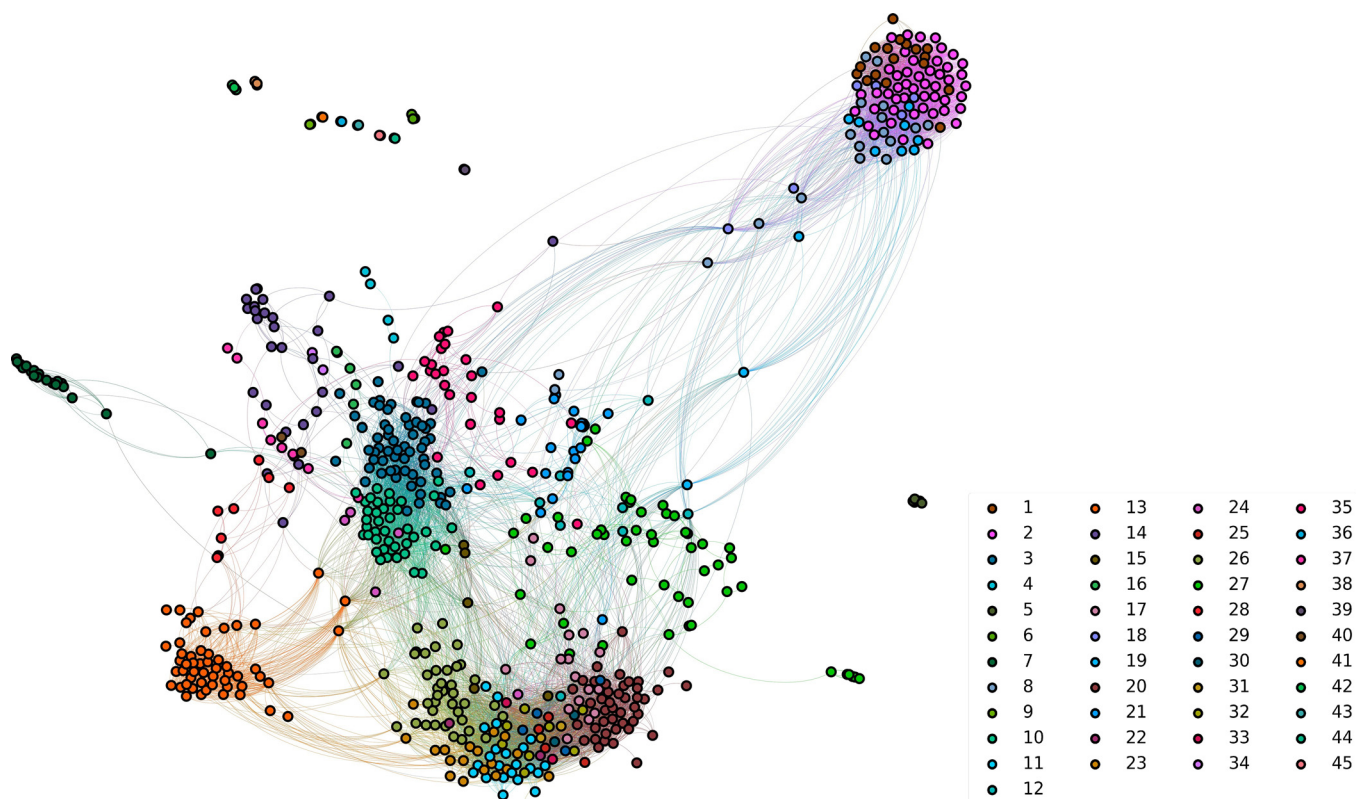


Figure 3. Similarity map of PubChem *in vitro* bioassays used in this study. Nodes represent PubChem assays; edges represent Jaccard dissimilarity computed based on chemical fragment bioprofiles. In total, 45 distinct clusters were identified by the Louvain modularity algorithm. Nodes belonging to the same cluster are presumed to have a shared biological relevance and were used to perform read-across.

have been shown to be highly correlated with acute toxicity in an earlier study (Zhang et al. 2014a). Cluster 2 (96% *ppv* in the cross-validation procedure) included not only cytotoxicity assays from NCI but also a significant number of bioassays with protein

targets regulating cell growth (e.g., steroid receptor coactivators 1 and 3, paternally expressed gene 3, hypoxia-inducible factor 2 alpha). Cluster 3 (87% *ppv* in the cross-validation procedure) contained 60 PubChem *in vitro* bioassays and was the largest among all

Table 1. Source depositors of the PubChem 640 assays used in this study.

Source institution	Number of assays	Cluster membership ^a
NCGC	200	2, 3, 4, 8, 9, 10, 12, 13, 14, 16, 17, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, 32, 35, 36, 37, 39, 40, 41, 42, 44, 45
Tox21	114	3, 9, 10, 11, 12, 15, 17, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 38, 44
DTP/NCI	96	1, 2, 5, 6, 7, 8, 10, 18, 19, 20, 21, 27
Scripps Research Institute Molecular Screening Center	67	2, 3, 4, 5, 8, 10, 14, 16, 18, 19, 20, 21, 27, 35, 37, 39, 40
Sanford-Burnham Center for Chemical Genomics	41	2, 3, 4, 5, 8, 10, 14, 15, 18, 20, 21, 27, 35, 37, 43, 45
Broad Institute	31	2, 3, 8, 10, 12, 14, 16, 18, 20, 21, 27, 34, 35
Cheminformatics and Chemogenomics Research Group	16	3, 15, 17, 20, 24, 27, 29, 32, 36
EPA DSSTox	11	3, 6, 7, 37, 38
Johns Hopkins Ion Channel Center	11	3, 10, 13, 17, 20
Southern Research Institute	10	3, 14, 15, 17, 20, 21, 24
Southern Research Specialized Biocontainment Screening Center	10	3, 12, 27, 37
Emory University Molecular Libraries Screening Center	9	3, 14, 21, 28
University of New Mexico	8	14, 18, 20, 21, 41
ICCB-Longwood/NSRB Screening Facility, Harvard Medical School	4	3, 12, 14
Vanderbilt High Throughput Screening Facility	3	3, 17, 20
Columbia University Molecular Screening Center	2	14
University of Pittsburgh Molecular Library Screening Center	2	3, 17
ChEMBL	2	14, 27
Milwaukee Institute for Drug Discovery	1	4
Institute for Research in Immunology and Cancer	1	10
Psychoactive Drug Screening Program	1	20

Note: ChEMBL, European Molecular Biology Laboratory chemistry database; DTP/NCI, Developmental Therapeutics Program/National Cancer Institute; EPA DSSTox, Environmental Protection Agency Distributed Structure-Searchable Toxicity Database; GPCR, G protein-coupled receptors; ICCB, Institute of Chemistry and Cell Biology; NCGC, National Center for Advancing Translational Sciences Chemical Genomics Center; NSRB, National Screening Laboratory for the Regional Centers of Excellence for Biodefense and Emerging Infectious Diseases; Tox21, Toxicity Testing in the 21st Century.

^aCluster membership displays an exhaustive list of all the clusters to which a particular source has at least one bioassay belong to, as identified by the Louvain modularity algorithm.

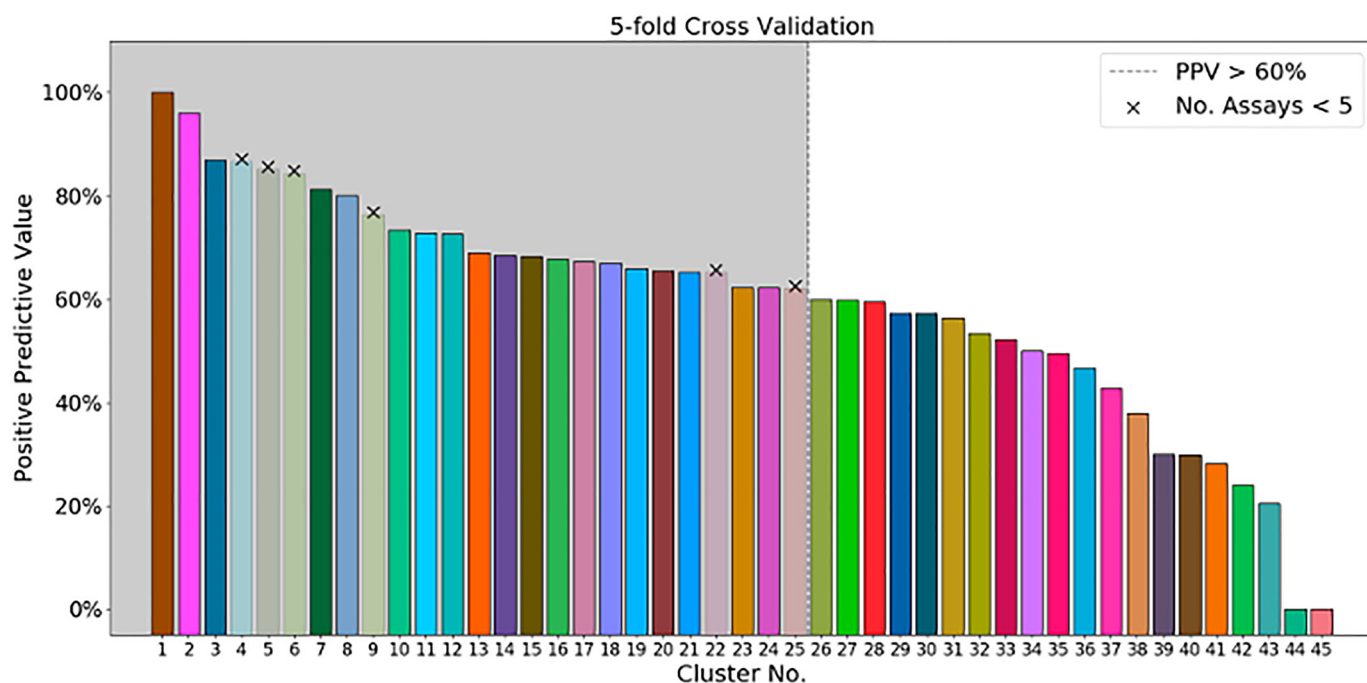


Figure 4. Individual cluster model-specific positive predictive value (*ppv*) results from fivefold cross-validation. Numbers along the x-axis correspond to clusters identified within the network in Figure 2. The first 25 clusters have *ppv* > 60% (shaded area), although five have less than five bioassays (marked by “x”) and were omitted from further analyses.

19 selected clusters. It consisted of data deposited from the National Center for Advancing Translational Sciences (formerly National Chemical Genomics Center) (25%), the Scripps Research Institute (20%), Tox21 (12%), Emory University Molecular Libraries Screening Center (10%), Johns Hopkins University (8%), and others (25%). Only a few of these bioassays were cytotoxicity assays, with the majority being associated with protein targets, including nuclear receptors, membrane channels, kinases, and various enzymes. Cluster 8 (*ppv* 80%) consisted of a mixture of cell viability, protein targets, and viral bioassays from five sources.

Ensemble Predictions for New Compounds

Prior experimentation has demonstrated that integration of multiple models can show superior predictivity to individual models (Kim et al. 2014; Solimeo et al. 2012; Sprague et al. 2014; Wang et al. 2015; Zhao et al. 2017; Zhu et al. 2009). In this project, one such approach was applied, which involved generating an ensemble model using the 19 models selected (i.e., those having *ppv* > 60%) to predict acute oral toxicity. For evaluation purposes, the resulted models were used to predict an external test set consisting of 639 compounds that were not present within the original training set. To build the ensemble model, bioprofile-based read-across prediction output values were averaged, per chemical, from all 19 models (Figure 5). Many new compounds contained varying amounts of *in vitro* biological data among the 19 PubChem bioassay clusters, and the read-across performance was also different among these 19 models. When applying increasing thresholds of confidence for the read-across predictions, some of these models were not able to provide predictions for new compounds due to insufficient data in the associated PubChem bioassays. To quantify this source of uncertainty, confidence values (*rcc*) were computed and evaluated. Ensemble model predictions made with a low confidence threshold (e.g., *rcc* values ≤ 10%, indicating that the external test set chemical and training set chemical were tested in few of the same bioassays) resulted in poor prediction of new compounds with *ppv* of 60%. However, when the confidence threshold was increased to eliminate unreliable

predictions (e.g., *rcc* values > 30%), the *ppv* value increased significantly (*ppv* > 76%). This improved ensemble model performance at higher confidence thresholds was similar to the cross-validation results of the most selected models shown in Figure 4.

Toxicity Pathways for Acute Oral Toxicity

To identify potential toxicity mechanisms from predictions, chemical fragment-*in vitro*-*in vivo* relationship analysis was performed within each individual cluster. Unfortunately, a complete dataset is critical for this analysis. To resolve the missing data issue, QSAR model predictions were used to fill in the missing data for the 3,842 compounds that were initially omitted from the training set due to insufficient bioassay data. The use of QSAR model predictions will add additional uncertainty since the QSAR model predictivity of PubChem bioassay responses for new compounds was generally around 70% (Figure S2).

The population of the bioprofile data by QSAR predictions can provide insights to reveal potential toxicity mechanisms within each cluster. For example, investigation into the chemical scaffolds of the predicted toxic compounds within a cluster can reveal toxicophores related to acute oral toxicity. These toxicophores can be used to explain toxicity mechanisms by integrating bioassay data used for read-across in relevant clusters. Two of the 19 clusters showed exceptional chemical fragment-*in vitro*-*in vivo* relationships.

Cluster 1, consisting of 17 NCI tumor cell line growth inhibition assays, had 24 predicted toxic compounds at an *rcc* of 100%, 13 of which were true positives (i.e., toxic *in vivo*) and 11 of which were false positives (i.e., nontoxic *in vivo* based on our toxic/nontoxic designations using a 1,000-mg/kg threshold). Examination of the chemical structures of the true positives and false positives (Figure 6; Excel Table S2) revealed distinctly different chemical scaffolds and suggested a potential toxicophore. Within the compounds predicted as toxic by this model, eight compounds had a steroid backbone (Figure 6B), and seven were classified as toxic in acute oral animal studies. Additionally, 10

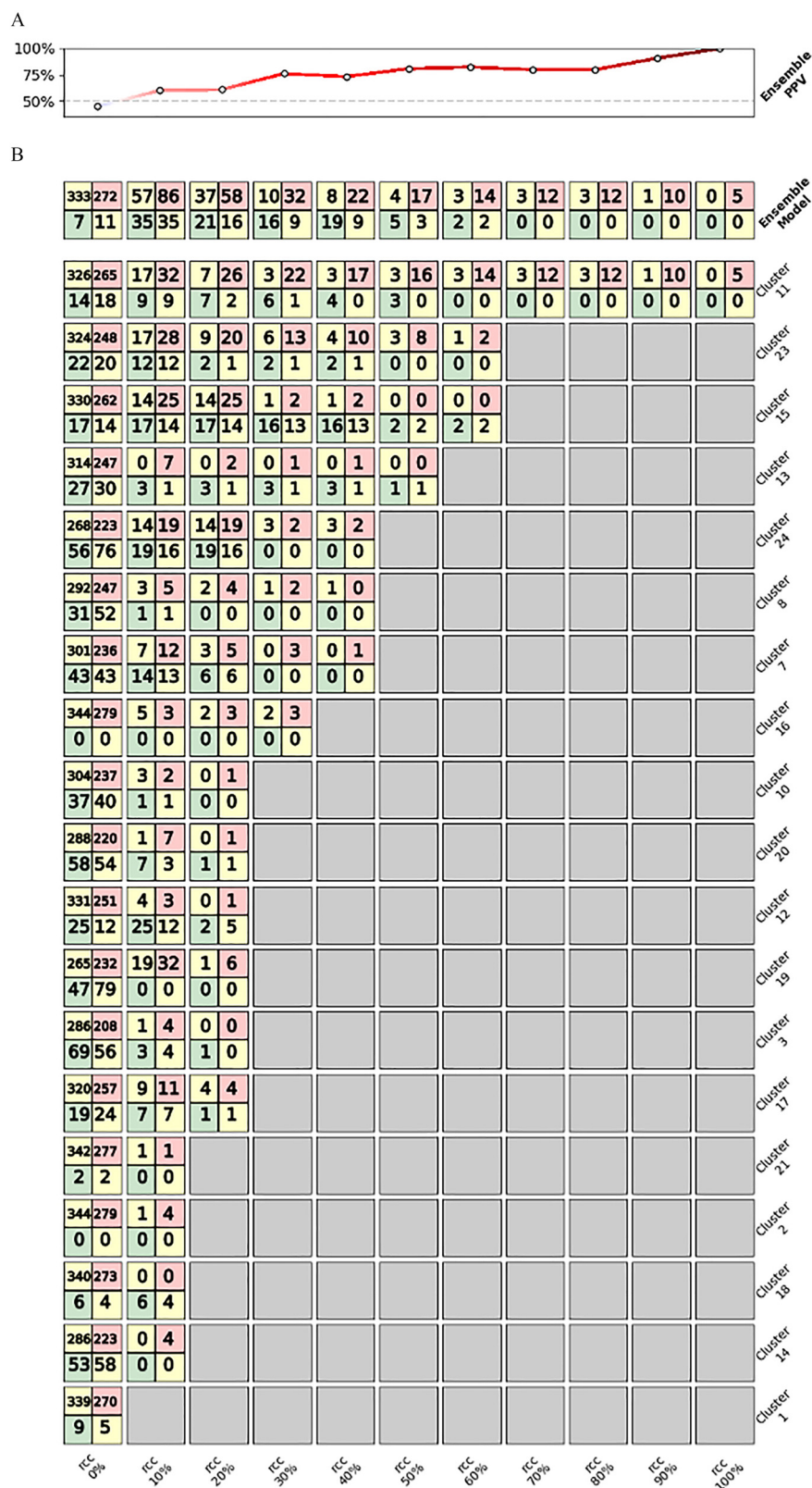


Figure 5. Bioprofile-based prediction performance on the external test set of 639 chemicals. The 639 external test set compounds were used to evaluate the predictivity of 19 clusters and the ensemble model at varying relative cluster confidence (rcc) levels. (A) Positive predictive value (ppv) of the ensemble model at increasing rcc thresholds (same x-axis as grid plot in panel B). (B) Grid plot showing confusion matrices of the ensemble model and the individual clusters at different rcc thresholds. Each confusion matrix contains the number of false positives (upper left corner), true positives (upper right corner), false negatives (lower left corner), and true negatives (lower right corner) for a model at a particular rcc threshold.

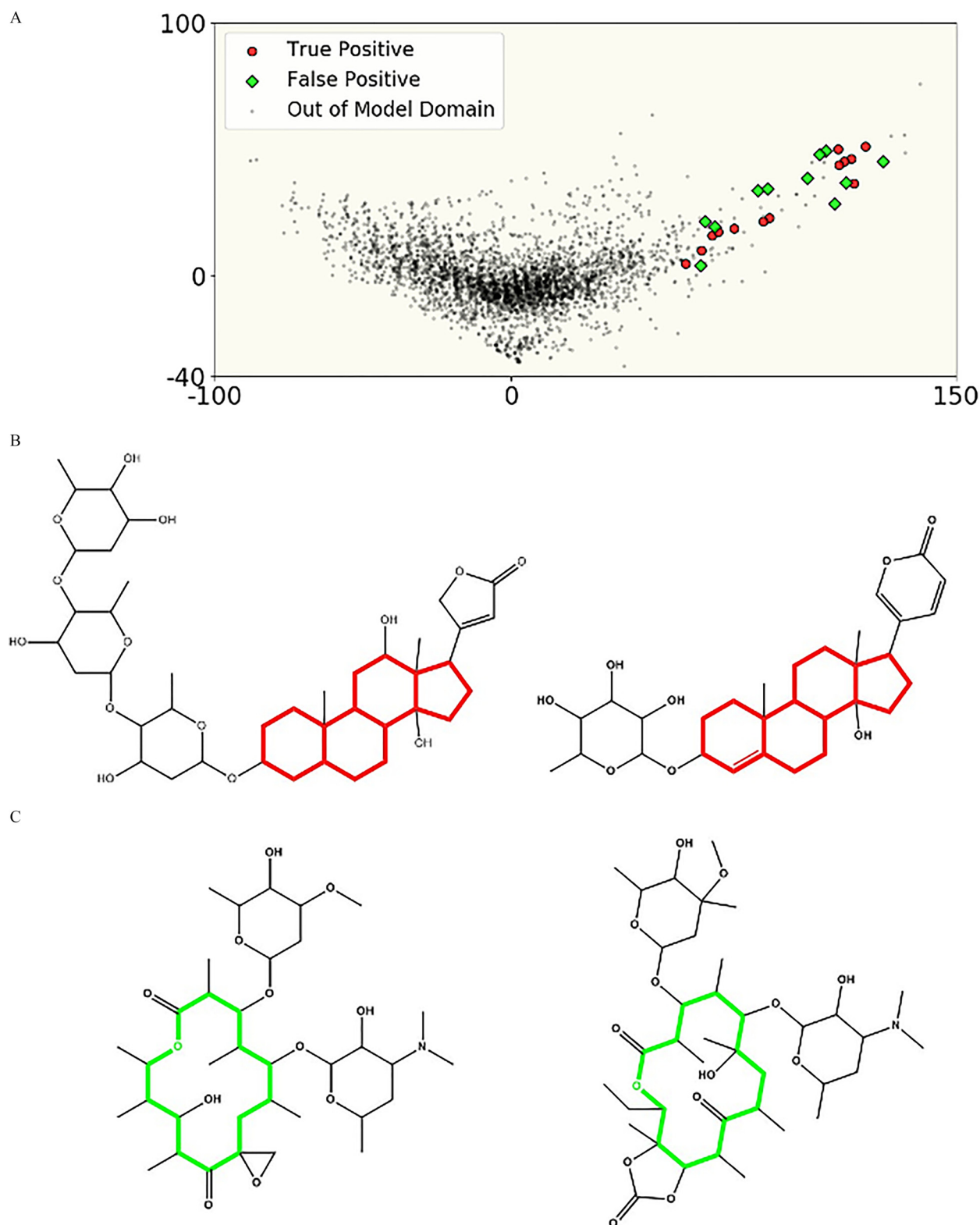


Figure 6. Cluster 1 predictions. (A) Principal component analysis (PCA) of Cluster 1 model predictions (true positives as highlighted circles and false positives as diamonds). (B) Two examples of true positive compounds with their common representative substructure in bold. (C) Two examples of false positive compounds with their common representative substructure in bold.

of the 17 bioassays in this model had this chemical fragment statistically relevant to their activity by Fisher's exact test.

Cluster 8 consisted of seven cytotoxicity assays and seven protein and viral targets, the latter of which are listed in Table 2. This

model showed a remarkable predictivity (modeling and external test sets had *ppv* of 80 and 63%, respectively). The associated toxicophore within this cluster is a trifluoromethyl-substituted benzimidazole ring (Figure 7). While this ring structure appears in 129 of

Table 2. PubChem Assays involving protein or viral targets within Cluster 8 model.

PubChem AID	Name	Target
488899	MITF measured in cell-based system using plate reader	Microphthalmia-associated transcription factor (<i>Homo sapiens</i>)
504444	Nrf2 qHTS screen for inhibitors	Nuclear factor erythroid 2–related factor 2 isoform 2 (<i>H. sapiens</i>)
540276	qHTS for inhibitors of binding or entry into cells for Marburg virus	Gene 4 small orf (Marburg virus)
588413	uHTS identification of Gli-Sufu antagonists in a luminescence reporter assay	Glioma-associated oncogene 1 (<i>Mus musculus</i>)
624169	Luminescence-based cell-based primary high- throughput screening assay to identify agonists of the mouse 5-hydroxy-tryptamine (serotonin) receptor 2A (HTR2A)	5-Hydroxytryptamine receptor 2A (<i>M. musculus</i>)
624354	uHTS identification of Caspase-8 TRAIL sensitizers in a luminescence assay	Tumor necrosis factor receptor superfamily member 10B isoform 1 precursor (<i>H. sapiens</i>)
651820	qHTS assay for inhibitors of hepatitis C virus (HCV)	Hepatitis C virus

Note: MITF, microphthalmia-associated transcription factor; Nrf2, nuclear factor erythroid 2–related factor 2; qHTS, quantitative high-throughput screening; TRAIL, tumor necrosis factor-related apoptosis-inducing ligand; uHTS, ultra-high-throughput screening.

the predicted chemicals, and all but one was toxic based on our classification criteria, not all were within the domain of this cluster's model. Yet the 46 toxic compounds that contained this fragment and were within the domain of this cluster's model were all correctly classified as toxic (Excel Table S3). This highlights the important role of an applicability domain assessment, which we integrated into the evaluation of the model. The compounds containing this toxicophore are pesticides and behave as casein kinase 2 (CK2) inhibitors, (e.g., 4,5,6,7-tetrabromobenzimidazole, CAS 749,234-11-5) (Adamson et al. 1984; Jones and Watson 1965; Pagano et al. 2004).

Discussion

We have recently advocated for the integration of biological information, such as data from PubChem, into read-across (Zhu

et al. 2016). In this study, we demonstrated a novel approach to integrate diverse *in vitro* data from a rapidly evolving public resource to serve as the foundation for bioprofile-based read-across predictions. The resulting bioassay cluster-based predictive models, and the ensemble model generated by combining cluster predictions, show great promise for predicting acute oral toxicity as well as informing on possible mechanisms contributing to toxicity.

The success of bioprofile-based read-across for the external test set (Figure 5) was different for each model as compared to the cross-validation within training sets (Figure 4). For example, Cluster 1, which mostly consisted of growth inhibition/cytotoxicity assays and yielded the best cross-validation accuracy (Figure 4), had no significant contribution to the external test set predictions, as shown in Figure 5. This is due to the lack of sufficient bioassay data for the external test set compounds in the subset of

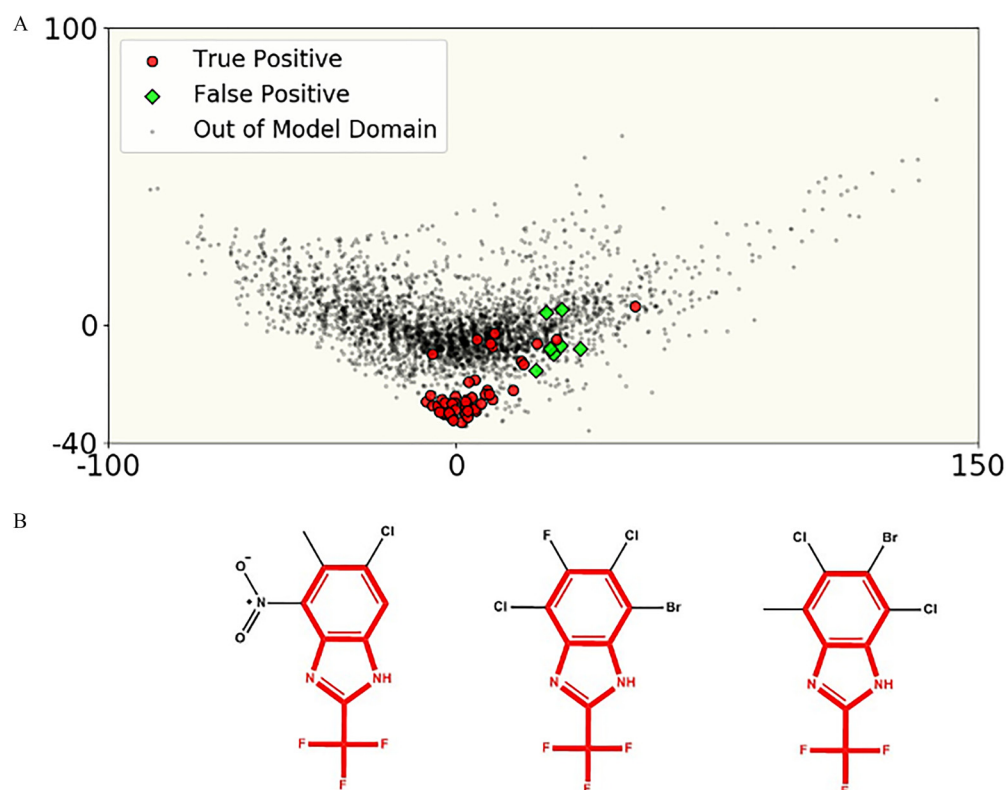


Figure 7. Cluster 8 predictions. (A) Principal component analysis (PCA) of Cluster 8 model predictions (true positives as highlighted circles and false positives as diamonds). (B) Three examples of true positives and the potential toxicophores associated with this cluster in bold are shown.

PubChem bioassays. Although cytotoxicity assays have been proven to be useful in predicting acute oral toxicity (Barile et al. 1994; Garle et al. 1994; Ukeliš et al. 2008), these bioassays will not contribute to predicting the toxicity of new compounds if the new compounds lack sufficient data. This issue can be solved by predicting new compound bioactivity *in vitro* (e.g., by QSAR modeling conducted herein and accepting an additional margin of uncertainty) or by testing new compounds using these bioassays. Consensus models have typically provided superior or equivalent predictivity relative to the individual models in our previous QSAR studies (Kim et al. 2014; Solimeo et al. 2012; Sprague et al. 2014; Wang et al. 2015; Zhao et al. 2017; Zhu et al. 2009). Here we used a similar approach, where an ensemble model was used to partially resolve this issue of insufficient data. Using this approach, the predictivity for new compounds in the external test set showed good *ppv* (76%) at reasonable prediction confidence (*rcc* of 30%) but with a low coverage (~10%). This reflects the importance of having enough bioassay data to evaluate new compounds for our bioprofile-based prediction approach to succeed.

In both fivefold cross-validation and the external evaluation, cytotoxicity assays appear as strong predictors of acute oral toxicity. The best performing models in fivefold cross-validation contain cytotoxicity assays from NCI (e.g., those in Clusters 1 and 2); however, they lacked sufficient data for evaluation purposes using the external test compounds. On the other hand, Clusters 11 and 23 showed accurate predictivity and contained many cytotoxicity assays from the Tox21 screening program. With reasonable confidence (*rcc* of 30%), they both showed high predictivity for the external test set compounds (*ppv* of 88% and 68%, respectively).

Unfortunately, the external test set compounds were not numerous enough to establish meaningful chemical fragment–*in vitro*–*in vivo* relationships. However, by using the QSAR predictions, enough data could be generated to see relationships between certain chemical fragments and *in vitro* bioassay activity responses within a cluster. This information can be used to enhance read-across predictions and/or explore toxicity mechanisms. While the use of QSAR undoubtedly introduces uncertainty into the read-across predictions, manual review of the model outputs (namely, chemical scaffolds of predicted compounds) can aid in bolstering confidence in the results. For example, identifying the distinction of true positive vs. false positive predictions within Cluster 1 (Figure 6) was explored. This cluster consisted entirely of cytotoxicity/growth inhibition bioassays in immortalized cell lines. Many of the true positives in these bioassays were steroids, which have been established as lead anticancer agents stemming from their affinity for nuclear receptors (Carvalho et al. 2010; Gupta et al. 2013). Indeed, the literature supports that steroid type toxicants induce acute toxicity because of cytotoxicity (Tantawy et al. 2017; Ur Rahman et al. 2017). On the other hand, 10 of the 11 false positive compounds were antibiotics that contain a large ring structure (greater than six members). Because of this unique chemical structure, antibiotics are likely to be misclassified by the QSAR models and should likely be excluded from the domain of applicability for the Cluster 1 model. These results suggest that cytotoxicity assays are reasonable alternative approaches for screening steroids to inform on acute oral toxicity and highlight the implications of conducting applicability domain assessments. Another consideration worth noting is that these bioassays use human cell lines but are being used to predict acute oral toxicity identified in rats. However, we do not expect a significant interspecies effect with regards to cytotoxicity assays because previous studies have demonstrated good concordance between cytotoxicity results obtained from rat and human cell lines (Clemenson and Ekwall 1999; Clothier et al. 2013).

Cluster 8 contained several pesticides that are putative CK2 inhibitors sharing a benzimidazole ring (Figure 7). CK2 is a

serine/threonine protein kinase targeting a wide array of proteins involved in several cell processes, including mediating cell cycle and apoptosis (Hamacher et al. 2007; Litchfield 2003; Yamane and Kinsella 2005), and at least two PubChem bioassays within this cluster are relevant to it. The first assay (PubChem AID 504444) screens for inhibitors of nuclear factor erythroid 2–related factor 2 (Nrf2), a transcription factor intimately involved in the cellular response to oxidative stress (Hur and Gray 2011). Nrf2 responds to oxidative stress by translocating to the nucleus, where it binds to the antioxidant response element (ARE) and induces expression of an array of genes encoding antioxidants (Jaiswal 2004). It has been shown that Nrf2 phosphorylation by CK2 is required for translocation to the nucleus and subsequent ARE activation (Apopa et al. 2008; Pi et al. 2007). Another assay (PubChem AID 588413) measures the inhibition of Gli1, a transcription factor involved in the hedgehog (Hh) signaling pathway (Ruiz i Altaba 1999; Villavicencio et al. 2000). The Hh pathway is involved in cell proliferation, cell maintenance, cell differentiation, and embryonic development (Gupta et al. 2010; Mahindroo et al. 2009). In addition to these processes, Hh signaling and Gli1 expression have been delineated as responders to oxidative stress, suggesting they have a role in regulating antioxidant genes (Chen et al. 2017; Yao et al. 2017). Similar to Nrf2, recent work has also shown CK2 to be a positive modulator of Gli1 activation (Jin et al. 2011) and downstream Hh pathway signaling (Zhang et al. 2012, 2014b). Thus, toxic compounds containing the toxiphore in Figure 7 may initiate a toxicity pathway by inhibition of CK2 and disruption of cell homeostasis in one of the following ways: *a*) by obstructing Nrf2 translocation to the nucleus, the antioxidant response is diminished and leaves the cell more susceptible to oxidative stress; and *b*) inhibition of Gli1 could result in silencing of the Hh pathway, disrupted cellular growth, and additional lowered antioxidant response. This example highlights the additional context that could be inferred from in-depth review of relationships between the bioassays within clusters to help inform on potential novel molecular mechanisms underlying acute oral toxicity.

Developing nonanimal models for acute oral toxicity evaluations still poses a significant challenge. While QSAR models built from the same rat acute oral toxicity training set offered an acceptable *ppv* on the same external test set compounds (*ppv* > 62%, shown in Excel Table S4), the prediction accuracy was lower than that achieved by the presented ensemble model (Figure 5). Furthermore, by leveraging *in vitro* bioactivity data, the bioprofile-based read-across method presented herein offers a solution to potential activity cliff issues existing in normal QSAR modeling (Maggiore 2006) because the new compound predictions are not limited to the information only obtained from chemical structures. To succeed, the integration of even more *in vitro* bioassays will likely be necessary to associate toxicity mechanisms for untested potential toxicants. Many of the toxicity mechanisms are obscure, so it is limiting to rely solely on the available bioassay data, especially unstructured public data, for toxicity evaluation purposes. This study shows that toxicity evaluations can be significantly enhanced by using meaningful biological data.

The new computational approach developed in this study can automatically identify useful biological data from public data sources and is capable of performing bioprofile-based read-across for untested compounds based on chemical fragment–*in vitro*–*in vivo* relationships. By doing so, chemical toxicity mechanisms can be elucidated. Such promising bioassays can then be used to characterize unknown substances. For example, these bioassays can be incorporated into weight of evidence approaches, such as an integrated testing strategy (Hartung et al. 2013; Rovida et al. 2015). Such strategies enable the incorporation of additional data that

may be critical for more relevant risk assessment, including bioassays capable of *in vitro* metabolism (Jacobs et al. 2013; McKim 2010; Yoon et al. 2012) as well as computational models to predict bioavailability (Bhatarai et al. 2015; Kim et al. 2014). Read-across studies using biological data strongly depend on the reliability of bioassay testing protocols and the quality of biological data. Potential experimental errors and other relevant issues (e.g., data reproducibility) may affect the confidence of data from bioassays. Although it is beyond the scope of this study, we have recommended potential solutions to improve the quality of public data in previous publications (Sedykh et al. 2011; Zhao et al. 2017).

Conclusions

For complex animal toxicity end points, such as acute oral toxicity, the complete replacement of animal testing is still not feasible. However, efforts to prioritize potentially hazardous chemicals by leveraging reliable and sufficient bioassay data that can be linked to specific toxicity mechanism(s) can significantly reduce the number of animals used, save great resources in chemical toxicology studies, and facilitate hazard assessment of high-priority chemicals. The data-driven profiling strategy presented in this study provides a novel way of extracting pertinent information from a daily updated, unstructured public resource. In contrast to previous studies, our method incorporates both chemical (i.e., chemical structure) and biological (*in vitro* bioassays) data into the workflow. This approach not only predicts acute oral toxicity classification but also infers biological mechanism information, offering novel insights into mechanisms of acute oral toxicity as well as *in vitro* bioassays and their utility for predicting *in vivo* toxicity. Furthermore, this method can easily be expanded to develop non-animal models to evaluate other complex animal toxicities beyond acute oral systemic toxicity.

Acknowledgments

D.P.R., W.W., and H.Z. were partially supported by the National Institute of Environmental Health Sciences (grant number R15ES023148), the Colgate-Palmolive Grant for Alternative Research, and the Johns Hopkins Center for Alternatives to Animal Testing (CAAT) grant. J.S. and A.K. were supported by NIH contract HHSN273201500010C to Integrated Laboratory Systems (ILS) in support of the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. The authors kindly acknowledge D. Allen and X. Chang of ILS for their help in preparation of the manuscript.

References

Adamson GW, Bawden D, Siggers DT. 1984. Quantitative structure-activity relationship studies of acute toxicity (LD50) in a large series of herbicidal benzimidazoles. *Pestic Sci* 15(1):31–39, <https://doi.org/10.1002/ps.2780150106>.

Allen TEH, Goodman JM, Gutsell S, Russell PJ. 2014. Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chem Res Toxicol* 27(12):2100–2112, PMID: 25354311, <https://doi.org/10.1021/tx500345j>.

Apopa PL, He X, Ma Q. 2008. Phosphorylation of Nrf2 in the transcription activation domain by casein kinase 2 (CK2) is critical for the nuclear translocation and transcription activation function of Nrf2 in IMR-32 neuroblastoma cells. *J Biochem Mol Toxicol* 22(1):63–76, PMID: 18273910, <https://doi.org/10.1002/jbt.20212>.

Attene-Ramos MS, Miller N, Huang R, Michael S, Itkin M, Kavlock RJ, et al. 2013. The Tox21 robotic platform for the assessment of environmental chemicals – from vision to reality. *Drug Discov Today* 18(15–16):716–723, PMID: 23732176, <https://doi.org/10.1016/j.drudis.2013.05.015>.

Barile FA, Dierickx PJ, Kristen U. 1994. *In vitro* cytotoxicity testing for prediction of acute human toxicity. *Cell Biol Toxicol* 10(3):155–162, PMID: 7994632, <https://doi.org/10.1007/BF00757558>.

Bhatarai B, Wilson DM, Bartels MJ, Chaudhuri S, Price PS, Carney EW. 2015. Acute toxicity prediction in multiple species by leveraging mechanistic ToxCast

mitochondrial inhibition data and simulation of oral bioavailability. *Toxicol Sci* 147(2):386–396, PMID: 26139166, <https://doi.org/10.1093/toxsci/kfv135>.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J Stat Mech Theory Mech* 2008(10):P10008, <https://doi.org/10.1088/1742-5468/2008/10/P10008>.

Breiman L. 2001. Random Forests. *Mach Learn* 45(1):5–32, <https://doi.org/10.1023/A:1010933404324>.

Browne P, Judson RS, Casey WM, Kleinstreuer NC, Thomas RS. 2015. Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ Sci Technol* 49(14):8804–8814, PMID: 26066997, <https://doi.org/10.1021/acs.est.5b02641>.

Carvalho JFS, Silva MMC, Moreira JN, Simões S, Sá e Melo ML. 2010. Sterols as anticancer agents: synthesis of ring-B oxygenated steroids, cytotoxic profile, and comprehensive SAR analysis. *J Med Chem* 53(21):7632–7638, PMID: 20931970, <https://doi.org/10.1021/jm1007769>.

Chen K-Y, Chiu C-H, Wang L-C. 2017. Anti-apoptotic effects of Sonic hedgehog signalling through oxidative stress reduction in astrocytes co-cultured with excretory-secretory products of larval *Angiostrongylus cantonensis*. *Sci Rep* 7:41574.

Clemenson C, Ekwall B. 1999. Overview of the final MEIC results: I. The *in vitro*–*in vitro* evaluation. *Toxicol In Vitro* 13(4–5):657–663, PMID: 20654531, [https://doi.org/10.1016/S0887-2333\(99\)00060-0](https://doi.org/10.1016/S0887-2333(99)00060-0).

Clothier R, Gómez-Lechón MJ, Kinsner-Ovaskainen A, Kopp-Schneider A, O'Connor JE, Prieto P, et al. 2013. Comparative analysis of eight cytotoxicity assays evaluated through the ACuteTox Project. *Toxicol In Vitro* 27(4):1347–1356, PMID: 22951948, <https://doi.org/10.1016/j.tiv.2012.08.015>.

Corvaro M, Gehen S, Andrews K, Chatfield R, Arasti C, Mehta J. 2016. GHS additivity formula: a true replacement method for acute systemic toxicity testing of agrochemical formulations. *Regul Toxicol Pharmacol* 82:99–110, PMID: 27765716, <https://doi.org/10.1016/j.yrtph.2016.10.007>.

Enoch SJ, Cronin MTD, Schultz TW, Madden JC. 2008. Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem Res Toxicol* 21(2):513–520, PMID: 18189367, <https://doi.org/10.1021/bx700322g>.

Friedman N, Geiger D, Goldszmidt M. 1997. Bayesian network classifiers. *Mach Learn* 29(2/3):131–163, <https://doi.org/10.1023/A:1007465528199>.

Garle MJ, Fentem JH, Fry JR. 1994. *In vitro* cytotoxicity tests for the prediction of acute toxicity *in vivo*. *Toxicol In Vitro* 8(6):1303–1312, PMID: 20693102, [https://doi.org/10.1016/0887-2333\(94\)90123-6](https://doi.org/10.1016/0887-2333(94)90123-6).

Gupta A, Kumar BS, Negi AS. 2013. Current status on development of steroids as anticancer agents. *J Steroid Biochem Mol Biol* 137:242–270, PMID: 23727548, <https://doi.org/10.1016/j.jsbmb.2013.05.011>.

Gupta S, Takebe N, LoRusso P. 2010. Targeting the Hedgehog pathway in cancer. *Ther Adv Med Oncol* 2(4):237–250, PMID: 21789137, <https://doi.org/10.1177/1758834010366430>.

Hamacher R, Saur D, Fritsch R, Reichert M, Schmid RM, Schneider G. 2007. Casein kinase II inhibition induces apoptosis in pancreatic cancer cells. *Oncol Rep* 18(3):695–701, PMID: 17671722, <https://doi.org/10.3892/or.18.3.695>.

Hartung T. 2016. Making big sense from big data in toxicology by read-across. *ALTEX* 33(2):83–93, PMID: 27032088, <https://doi.org/10.14573/altex.1603091>.

Hartung T. 2017. Evolution of toxicological science: the need for change. *IJRAM* 20(1/2/3):21–45, <https://doi.org/10.1504/IJRAM.2017.082570>.

Hartung T, Luechtefeld T, Maertens A, Kleensang A. 2013. Integrated testing strategies for safety assessments. *ALTEX* 30(1):3–18, PMID: 23338803, <https://doi.org/10.14573/altex.2013.1.003>.

Hartung T, Rovida C. 2009. Chemical regulators have overreached. *Nature* 460(7259):1080–1081, PMID: 19713914, <https://doi.org/10.1038/4601080a>.

Hewitt M, Ellison CM, Enoch SJ, Madden JC, Cronin M. 2010. Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reprod Toxicol* 30(1):147–160, PMID: 20006701, <https://doi.org/10.1016/j.reprotox.2009.12.003>.

Hur W, Gray NS. 2011. Small molecule modulators of antioxidant response pathway. *Curr Opin Chem Biol* 15(1):162–173, PMID: 21195017, <https://doi.org/10.1016/j.cbpa.2010.12.009>.

Jacobs MN, Laws SC, Willett K, Schmieder P, Odum J, Bovee TF. 2013. *In vitro* metabolism and bioavailability tests for endocrine active substances: what is needed next for regulatory purposes? *ALTEX* 30(3):331–351, PMID: 23861078, <https://doi.org/10.14573/altex.2013.3.331>.

Jaiswal AK. 2004. Nrf2 signaling in coordinated activation of antioxidant gene expression. *Free Radic Biol Med* 36(10):1199–1207, PMID: 15110384, <https://doi.org/10.1016/j.freeradbiomed.2004.02.074>.

Jin Z, Mei W, Strack S, Jia J, Yang J. 2011. The antagonistic action of B56-containing protein phosphatase 2As and casein kinase 2 controls the phosphorylation and Gli turnover function of Daz interacting protein 1. *J Biol Chem* 286(42):36171–36179, PMID: 21878643, <https://doi.org/10.1074/jbc.M111.274761>.

- Jones OTG, Watson WA. 1965. Activity of 2-trifluoromethylbenzimidazoles as uncouplers of oxidative phosphorylation. *Nature* 208(5016):1169–1170, PMID: 5870313, <https://doi.org/10.1038/2081169a0>.
- Judson RS, Maggintay FM, Chickarmar V, Haskell C, Tania N, Taylor J, et al. 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol Sci* 148(1):137–154, PMID: 26272952, <https://doi.org/10.1093/toxsci/kfv168>.
- Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect* 117(5):685–695, PMID: 19479008, <https://doi.org/10.1289/ehp.0800168>.
- Kim MT, Huang R, Sedykh A, Wang W, Xia M, Zhu H. 2016. Mechanism profiling of hepatotoxicity caused by oxidative stress using antioxidant response element reporter gene assay models and big data. *Environ Health Perspect* 124(5):634–641, PMID: 26383846, <https://doi.org/10.1289/ehp.1509763>.
- Kim MT, Sedykh A, Chakravarti SK, Saiakhov RD, Zhu H. 2014. Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. *Pharm Res* 31(4):1002–1014, PMID: 24306326, <https://doi.org/10.1007/s11095-013-1222-1>.
- Kim S, Thiessen PA, Bolton EE, Bryant SH. 2015. PUG-SOAP and PUG-REST: web services for programmatic access to chemical information in PubChem. *Nucleic Acids Res* 43(W1):W605–W611, PMID: 25934803, <https://doi.org/10.1093/nar/gkv396>.
- Kinsner-Ovaskainen A, Bulgheroni A, Hartung T, Prieto P. 2009. ECVAM's ongoing activities in the area of acute oral toxicity. *Toxicol In Vitro* 23(8):1535–1540, PMID: 19591916, <https://doi.org/10.1016/j.tiv.2009.07.004>.
- Kleinstreuer NC, Ceger P, Watt ED, Martin M, Houck K, Browne P, et al. 2017. Development and validation of a computational model for androgen receptor activity. *Chem Res Toxicol* 30(4):946–964.
- Koleva YK, Madden JC, Cronin M. 2008. Formation of Categories from structure–activity relationships to allow read-across for risk assessment: toxicity of α,β -unsaturated carbonyl compounds. *Chem Res Toxicol* 21(12):2300–2312, PMID: 19053326, <https://doi.org/10.1021/tx8002438>.
- Litchfield DW. 2003. Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem J* 369(Pt 1):1–15, PMID: 12396231, <https://doi.org/10.1042/bj20021469>.
- Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, et al. 2013. Integrative chemical–biological read-across approach for chemical hazard classification. *Chem Res Toxicol* 26(8):1199–1208, PMID: 23848138, <https://doi.org/10.1021/tx400110f>.
- Luechtefeld T, Maertens A, Russo DP, Rovida C, Zhu H, Hartung T. 2016. Analysis of public oral toxicity data from REACH registrations 2008–2014. *ALTEX* 33(2):111–122.
- Maggiore GM. 2006. On outliers and activity cliffs—why QSAR often disappoints. *J Chem Inf Model* 46(4):1535, PMID: 16859285, <https://doi.org/10.1021/ci060117s>.
- Mahindroo N, Punchihewa C, Fujii N. 2009. Hedgehog-Gli signaling pathway inhibitors as anticancer agents. *J Med Chem* 52(13):3829–3845, PMID: 19309080, <https://doi.org/10.1021/jm801420y>.
- McKim JM Jr. 2010. Building a tiered approach to in vitro predictive toxicity screening: a focus on assays with in vivo relevance. *Comb Chem High Throughput Screen* 13(2):188–206, PMID: 20053163, <https://doi.org/10.2174/138620710790596736>.
- National Research Council. 2007. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press.
- Pagano MA, Andrzejewska M, Ruzzene M, Sarno S, Cesaro L, Bain J, et al. 2004. Optimization of protein kinase CK2 inhibitors derived from 4,5,6,7-tetrabromo-benzimidazole. *J Med Chem* 47(25):6239–6247, PMID: 15566294, <https://doi.org/10.1021/jm049854a>.
- Patlewicz G, Ball N, Becker RA, Booth ED, Cronin MTD, Kroese D, et al. 2014. Read-across approaches - misconceptions, promises and challenges ahead. *ALTEX* 31(4):387–396, PMID: 25368965, <https://doi.org/10.14573/altex.1410071>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. 2011. Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830.
- Pi J, Bai Y, Reece JM, Williams J, Liu D, Freeman ML, et al. 2007. Molecular mechanism of human Nrf2 activation and degradation: role of sequential phosphorylation by protein kinase CK2. *Free Radic Biol Med* 42(12):1797–1806, PMID: 17512459, <https://doi.org/10.1016/j.freeradbiomed.2007.03.001>.
- Rogers D, Hahn M. 2010. Extended-connectivity fingerprints. *J Chem Inf Model* 50(5):742–754, PMID: 20426451, <https://doi.org/10.1021/ci100050t>.
- Rovida C, Alépée N, Api AM, Baskett DA, Bois FY, Caloni F, et al. 2015. Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32(1):25–40, PMID: 25413849, <https://doi.org/10.14573/altex.1411011>.
- Ruiz i Altaba A. 1999. Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Dev Camb Engl* 126(14):3205–3216.
- Russo DP, Kim MT, Wang W, Pinolini D, Shende S, Strickland J, et al. 2017. CIIPro: a new read-across portal to fill data gaps using public large-scale chemical and biological data. *Bioinformatics* 33(3):464–466, PMID: 28172359, <https://doi.org/10.1093/bioinformatics/btw640>.
- Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, et al. 2011. Use of in vitro HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. *Environ Health Perspect* 119(3):364–370, PMID: 20980217, <https://doi.org/10.1289/ehp.1002476>.
- Solimeo R, Zhang J, Kim M, Sedykh A, Zhu H. 2012. Predicting chemical ocular toxicity using a combinatorial QSAR approach. *Chem Res Toxicol* 25(12):2763–2769, PMID: 23148656, <https://doi.org/10.1021/tx300393v>.
- Sprague B, Shi Q, Kim MT, Zhang L, Sedykh A, Ichiishi E, et al. 2014. Design, synthesis and experimental validation of novel potential chemopreventive agents using random forest and support vector machine binary classifiers. *J Comput Aided Mol Des* 28(6):631–646, PMID: 24840854, <https://doi.org/10.1007/s10822-014-9748-9>.
- Strickland J, Clippinger AJ, Brown J, Allen D, Jacobs A, Matheson J, et al. 2018. Status of acute toxicity testing requirements and data uses by U.S. regulatory agencies. *Regul Toxicol Pharmacol* 94:183–196.
- Tantawy MA, Nafie MS, Elmegeed GA, Ali IAI. 2017. Auspicious role of the steroidal heterocyclic derivatives as a platform for anti-cancer drugs. *Bioorganic Chem* 73:128–146, PMID: 28668650, <https://doi.org/10.1016/j.bioorg.2017.06.006>.
- U.S. EPA (Environmental Protection Agency). 2012. *Label Review Manual*. Washington, DC: U.S. EPA.
- U.S. EPA. 2016. Pesticides: Freedom of Information Act (FOIA). Updated 20 February 2016. https://archive.epa.gov/pesticides/chemicalsearch/chemical/foia/web/html/reading_room.html. [accessed 1 March 2010].
- Ukelis U, Kramer P-J, Olejniczak K, Mueller SO. 2008. Replacement of in vivo acute oral toxicity studies by in vitro cytotoxicity methods: opportunities, limits and regulatory status. *Regul Toxicol Pharmacol* 51(1):108–118, PMID: 18362045, <https://doi.org/10.1016/j.yrtph.2008.02.002>.
- Ur Rahman S, Ismail M, Khurram M, Ullah I, Rabbi F, Iriti M. 2017. Bioactive steroids and saponins of the genus *Trillium*. *Mol Basel Switz* 22(12):E2156, PMID: 29206216, <https://doi.org/10.3390/molecules221221>.
- Villavicencio EH, Walterhouse DO, Iannaccone PM. 2000. The sonic hedgehog-patched-gli pathway in human development and disease. *Am J Hum Genet* 67(5):1047–1054, PMID: 11001584, [https://doi.org/10.1016/S0002-9297\(07\)62934-6](https://doi.org/10.1016/S0002-9297(07)62934-6).
- Walum E. 1998. Acute oral toxicity. *Environ Health Perspect* 106(suppl 2):497–503, <https://doi.org/10.2307/3433801>.
- Wang NCY, Jay Zhao Q, Wesselkamper SC, Lambert JC, Petersen D, Hess-Wilson JK. 2012. Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach. *Regul Toxicol Pharmacol* 63(1):10–19, PMID: 22369873, <https://doi.org/10.1016/j.yrtph.2012.02.006>.
- Wang W, Kim MT, Sedykh A, Zhu H. 2015. Developing enhanced blood–brain barrier permeability models: integrating external bio-assay data in QSAR modeling. *Pharm Res* 32(9):3055–3065, PMID: 25862462, <https://doi.org/10.1007/s11095-015-1687-1>.
- Wu S, Fisher J, Naciff J, Laufersweiler M, Lester C, Daston G, et al. 2013. Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem Res Toxicol* 26(12):1840–1861, PMID: 24206190, <https://doi.org/10.1021/tx400226u>.
- Yamane K, Kinsella TJ. 2005. Casein kinase 2 regulates both apoptosis and the cell cycle following DNA damage induced by 6-thioguanine. *Clin Cancer Res* 11(6):2355–2363, PMID: 15788687, <https://doi.org/10.1158/1078-0432.CCR-04-1734>.
- Yang C, Tarkhov A, Marusczyk J, Bienfait B, Gasteiger J, Kleinoeder T, et al. 2015. New publicly available chemical query language, CSRML, to support chemistry representations for application to data mining and modeling. *J Chem Inf Model* 55(3):510–528, PMID: 25647539, <https://doi.org/10.1021/ci500667v>.
- Yao PJ, Manor U, Petralia RS, Brose RD, Wu RTY, Ott C, et al. 2017. Sonic hedgehog pathway activation increases mitochondrial abundance and activity in hippocampal neurons. *Mol Biol Cell* 28(3):387–395, PMID: 27932496, <https://doi.org/10.1091/mbc.e16-07-0553>.
- Yoon M, Campbell JL, Andersen ME, Clewell HJ. 2012. Quantitative in vitro to in vivo extrapolation of cell-based toxicity assay results. *Crit Rev Toxicol* 42(8):633–652, PMID: 22667820, <https://doi.org/10.3109/10408444.2012.692115>.
- Zhang J, Hsieh JH, Zhu H. 2014a. Profiling animal toxicants by automatically mining public bioassay data: a big data approach for computational toxicology. *PLoS One* 9(6):e99863, PMID: 24950175, <https://doi.org/10.1371/journal.pone.0099863>.
- Zhang S, Wang Y, Mao JH, Hsieh D, Kim IJ, Hu LM, et al. 2012. Inhibition of CK2 α down-regulates Hedgehog/Gli signaling leading to a reduction of a stem-like side population in human lung cancer cells. *PLoS One* 7(6):e38996, PMID: 22768056, <https://doi.org/10.1371/journal.pone.0038996>.
- Zhang S, Yang YL, Wang Y, You B, Dai Y, Chan G, et al. 2014b. CK2 α , over-expressed in human malignant pleural mesothelioma, regulates the Hedgehog

- signaling pathway in mesothelioma cells. *J Exp Clin Cancer Res* 33:93, PMID: [25422081](https://pubmed.ncbi.nlm.nih.gov/25422081/), <https://doi.org/10.1186/s13046-014-0093-6>.
- Zhao L, Wang W, Sedykh A, Zhu H. 2017. Experimental errors in QSAR modeling sets: what we can do and what we cannot do. *ACS Omega* 2(6):2805–2812, PMID: [28691113](https://pubmed.ncbi.nlm.nih.gov/28691113/), <https://doi.org/10.1021/acsomega.7b00274>.
- Zhu H, Bouhifd M, Donley E, Egnash L, Kleinstreuer N, Kroese ED, et al. 2016. Supporting read-across using biological data. *ALTEX* 33(2):167–182, PMID: [26863516](https://pubmed.ncbi.nlm.nih.gov/26863516/), <https://doi.org/10.14573/altex.1601252>.
- Zhu H, Martin TM, Ye L, Sedykh A, Young DM, Tropsha A. 2009. Quantitative structure–activity relationship modeling of rat acute toxicity by oral exposure. *Chem Res Toxicol* 22(12):1913–1921, PMID: [19845371](https://pubmed.ncbi.nlm.nih.gov/19845371/), <https://doi.org/10.1021/tx900189p>.
- Zhu H, Zhang J, Kim MT, Boison A, Sedykh A, Moran K. 2014. Big data in chemical toxicity research: the use of high-throughput screening assays to identify potential toxicants. *Chem Res Toxicol* 27(10):1643–1651, PMID: [25195622](https://pubmed.ncbi.nlm.nih.gov/25195622/), <https://doi.org/10.1021/tx500145h>.